

Table S3: Recapture fractions of other datasets by the GPC (at 99% similarity).

taxon	EMP	SILVA	UBA	IMG/M	RDP	GTDB
Bacteria	0.82	0.91	0.80	0.88	0.80	0.90
Archaea	0.62	0.74	0.86	NA	0.79	0.66